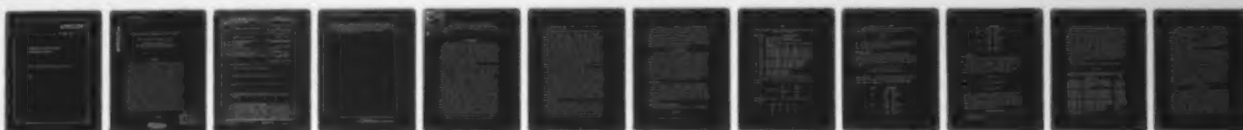
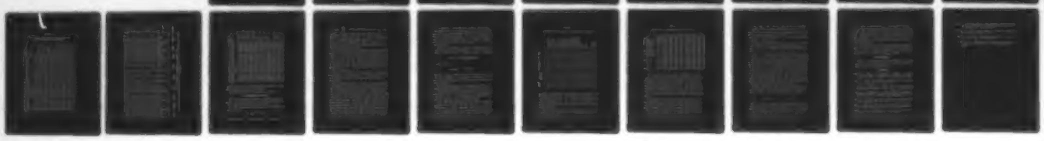
AD-A035 739 NORTH CAROLINA UNIV AT CHAPEL HILL DEPT OF STATISTICS F/G 12/1
A MONTE-CARLO STUDY OF ROBUST ESTIMATORS OF LOCATION.(U)
1976 E J WEGMAN, R J CARROLL AF-AFOSR-2840-75
UNCLASSIFIED AFOSR-TR-77-0022 NL

1 OF 1
AD-A
035 739





END
DATE
FILMED
3-25-77
NTIS

U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

AD-A035 739

A MONTE-CARLO STUDY OF ROBUST ESTIMATORS OF LOCATION

NORTH CAROLINA UNIVERSITY AT CHAPEL HILL

1976

A MONTE-CARLO STUDY OF ROBUST ESTIMATORS OF LOCATION

Edward J. Wegman and Raymond J. Carroll

Department of Statistics
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina 27514

ABSTRACT

Andrews et al (1972) carried out an extensive Monte Carlo study of robust estimators of location. Their conclusions were that the hampel and the skipped estimates, as classes, seemed to be preferable to some of the other currently fashionable estimators. The present study extends this work to include estimators not previously examined. The estimators are compared over short-tailed as well as long-tailed alternatives and also over some dependent data generated by first-order autoregressive schemes. The conclusions of the present study are threefold. First, from our limited study, none of the so-called robust estimators are very robust over short-tailed situations. More work seems to be necessary in this situation. Second, none of the estimators perform very well in dependent data situations, particularly when the correlation is large and positive. This seems to be a rather pressing problem. Finally, for long-tailed alternatives, the hampel estimators and Hogg-type adaptive versions of the hampels are the strongest classes. The adaptive hampels neither uniformly outperform nor are they outperformed by the hampels. However, the

1976

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR - TR - 77 - 0022	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A MONTE-CARLO STUDY OF ROBUST ESTIMATORS OF LOCATION		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Edward J. Wegman and R. J. Carroll		8. CONTRACT OR GRANT NUMBER(s) AFOSR 75-2840
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of North Carolina Department of Statistics Chapel Hill, NC 27514		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332		12. REPORT DATE 1976
		13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Robst estimation, Monte Carlo, Time Series, Adaptive Estimation, Trimmed means, m-estimates

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The present study extends the work of Andrews et al (1972) to include estimators not previously examined. They are compared over short-tailed as well as long tailed alternatives and also over some dependent data generated by first order autoregressive schemes. The conclusions of this study are: none of the so-called robust estimators are very robust over short-tailed situations; none of the estimators perform very well in dependent data situations, particularly when the correlation is large and positive; for long-tailed alternatives, the hampel estimators and Hogg-type adaptive versions

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20 Abstract

of the hampels are the strongest classes. The adaptive hampels neither uniformly outperform nor are they outperformed by the hampels. However, the superiority in terms of maximum relative efficiency goes to the adaptive hampels. That is the adaptive hampels, under their worst performance, are superior to the usual hampels, under their worst performance.

superiority in terms of maximum relative efficiency goes to the adaptive hampels. That is the adaptive hampels, under their worst performance, are superior to the usual hampels, under their worst performance.

1. INTRODUCTION

Andrews et al (1972) have published a very extensive and informative Monte-Carlo study of robust estimation of location in a symmetric probability density. This study involved some 68 estimators of location as well as 14 distinct sampling distributions. Sample sizes used were 5, 10, 20 and 40 although for most of the sampling situations only the sample size of 20 was investigated. For the estimators and sampling situations examined, they provide a very complete and satisfactory picture.

We are motivated, however, to supplement this study for three basic reasons. In Andrews et al, *(Princeton Robustness Study)* (1972, p. 67), hereafter referred to as *PRS*, short-tailed sampling situations are ruled out of consideration with the statement, "Robustness for short-tailed distributions was thought to be a rather special case, arising in situations that are usually rather easily recognized in practice." It is our contention that short-tailed data do arise in practice. For example, Professor R. V. Hogg points out that studies at the Iowa Testing Service, home of the ACT college entrance examination, clearly show that scores on these examinations tend to arise from short-tailed distributions. Our personal experience with instrumentation data from U. S. Naval projects also convinces us that data passing through one or more sets of electronic instruments often tend to be characterized by a short-tailed distribution. Moreover, since some of the location estimators were designed to protect against short-tailed alternatives as well as long-tailed possibilities, we believe that examination of long-tailed alternatives alone faults such estimators unfairly, rather akin to discovering a two-sided test is not most powerful against one-sided alternatives. Even ignoring this

aspect, the question remains, "What are desirable estimators given short-tailed alternatives?"

A second motivation was provided by the rather dismal situation in the time series context. The sample mean is routinely shown to be a consistent estimator of the "center" of a stationary time series with the clear implication that a time series is centered by subtracting the sample mean. One might suggest that one first remove the autocorrelation by fitting some sort of autoregressive (AR) or moving average (MA) scheme and then use $\bar{X}$ to estimate the center of the errors which in the stationary case is the same as the center of the time series. Unfortunately to fit an AR or a MA scheme requires that we have a zero mean time series, i.e. a centered time series. Thus a robust estimate of location in a time series context is a matter of practical concern. If the sample mean is unsatisfactory for independent, but non-normal data, how much worse must it be for correlated data?

Our third motivation arises from a personal conviction that adaptive estimators, properly formulated, ought to be very successful. In particular, we observe that because of the contributions of Professors Hampel and Huber to the *PRS*, the so-called hampel and closely related M-estimators were extensively studied. At least 25 of the 68 estimators studied involve either a hampel or a M-estimator. The M-estimators and particularly the hampels perform very well both because they are good estimators and because they were fine-tuned. The adaptive estimators, however, were not similarly fine-tuned and do not fare as well in the final analysis. To state our conviction succinctly, if hampels are good, adaptive hampels should be better.

The paper is divided into five parts. Section 2 lists the estimators studied in this paper and presents a short discussion. Section 3 discusses the details of the Monte-Carlo aspects of this work while section 4 contains the tables of results. Finally, section 5 contains our reactions to and conclusions about the results. We note here that this paper is not intended to compete in scope or

in detail with the *PRS*. The latter was a long and arduous study. We believe because of this, the *PRS* has and will formulate directions of research for some time, unfortunately in our view, away from adaptive estimators. Our intention is, therefore, to present evidence that adaptive estimators are at least competitive. Our adaptive procedures are based on experimentation and intuition and should be taken as first generation refinements. We are quite sure adaptive estimators can be further improved.

2. THE ESTIMATORS

In choosing which estimators to examine in this study, we were guided by the results of the *PRS*. As standards of comparison, we chose M-estimators and the related hampels and the trimmed means. Against these standards, we measured various forms of adaptive estimators and several miscellaneous estimators – including the Hodges-Lehmann, Normal Scores Rank Estimator, Johns', and an estimator based on skipping. We follow the routine established in the *PRS* by listing the estimators together with the mnemonic codes in table I. We note here that the codes described below agree with those in the *PRS* when an estimator is common to both works.

Trimmed Means

A simple scheme for robustifying the mean is to eliminate "extreme" observations. The $\alpha(100)\%$ symmetric trimmed mean discards the $[(N+1)\alpha]$ largest and $[(N+1)\alpha]$ smallest observations and computes the mean of the remaining observations. Here, $[\cdot]$ is the greatest integer function. The outer-mean, OM, is sometimes used in short-tailed situations and is computed as the mean of the trimmings with $\alpha = .25$.

Huber's M-Estimators and Hampels

M-estimators of location are solutions, T, of the equation

$$\sum_{j=1}^{n} \psi\left(\frac{X_j - T}{S}\right) = 0.$$

Hampel (1974) proposed estimating the scale, S, with

TABLE I

A Brief Description of the 19 Estimators of Location Together With a Mnemonic Code for Each

Number	Code	Short Description
1	10%	10% symmetrically trimmed mean
2	50%	50% symmetrically trimmed mean (median)
3	M	Mean
4	OM	Outer mean, mean of trimming after 25% symmetrical trimming
5	12A	One-step hampel M-estimate, ψ bends at 1.2,3.5,8.0
6	17A	One-step hampel M-estimate, ψ bends at 1.7,3.4,8.5
7	21A	One-step hampel M-estimate, ψ bends at 2.1,4.0,8.2
8	25A	One-step hampel M-estimate, ψ bends at 2.5,4,5,9.5
9	ADA	Adaptive hampel M-estimate, ψ bends at ADA,4.5,8.0
10	HG1	Hogg-type adaptor using trimmed means 38%,19%,M,OM
11	HG2	Hogg-type adaptor using trimmed means 38%,25%,10%
12	1.81*A	Hogg-type adaptor using hampels 25A,21A,12A
13	1.90A	Hogg-type adaptor using hampels 25A,ADA,17A
14	1.95A	Hogg-type adaptor using hampels 25A,ADA,17A
15	2.00A	Hogg-type adaptor using hampels 21A,12A
16	H/L	Hodges-Lehmann estimator
17	RN	Normal scores rank estimator
18	JOH	Johns' adaptive estimator
19	5T4	Multiply-skipped mean, max(5k,2)≤.6N deleted

$\text{med}|x_i - 50\%|/.6745$. The hampel estimators are M-estimators with ψ given by

$$\psi(x) = \text{sgn}\, x \cdot \begin{cases} |x| & 0 \le |x| < a \\ a & a \le |x| < b \\ \dfrac{c-|x|}{c-b} \cdot a & b \le |x| < c \\ 0 & |x| \ge c\ . \end{cases}$$

The parameters a, b and c are given below together with the code symbol.

Code	a	b	c
12A	1.2	3.5	8.0
17A	1.7	3.4	8.5
21A	2.1	4.0	8.2
25A	2.5	4.5	9.5

ADA is a mildly adaptive form of the hampel.

The hampels examined in this work are one-stop estimators (cf. Bickel (1975)), that is they are not the exact root of

$$\sum_{i=1}^{n} \psi\left(\frac{X_i - T}{S}\right) = 0,$$

but the result of one iteration of the Newton-Raphson method using 50% (the median) as starting value. More complete details of these estimators may be found in the *PRS*.

Hogg-Type Adaptors

Hogg (1974) has suggested an adaptation procedure which chooses among two or more estimates of the center depending on the value of some statistic chosen to measure tail length. Hogg originally suggested use of the kurtosis, while more recent suggestions include

$$Q_1 = \{U(.05)-L(.05)\}/\{U(.50)-L(.50)\}$$
$$Q_2 = \{U(.20)-L(.20)\}/\{U(.50)-L(.50)\},$$

where $U(\alpha)\{L(\alpha)\}$ is the mean of the largest (smallest) $[(N+1)\alpha]$ observations. The exact form of the various adaptors is given below. Here k% denotes a k% symmetrically trimmed mean.

Code	Formulation
HG1	$T = \begin{cases} 38\% & Q_1 > 3.2 \\ 19\% & 2.6 < Q_1 \le 3.2 \\ M & 2.0 < Q_1 \le 2.6 \\ OM & Q_1 \le 2.0 \end{cases}$
HG2	$T = \begin{cases} 38\% & Q_2 > 1.87 \\ 25\% & 1.81 < Q_2 \le 1.87 \\ 10\% & Q_2 \le 1.81 \end{cases}$
1.81*A	$T = \begin{cases} 21A & 1.81 < Q_2 \le 1.87 \\ 25A & Q_2 \le 1.81 \\ 12A & Q_2 > 1.87 \end{cases}$

Code	Formulation
1.90A	$T = \begin{cases} ADA & 1.90<Q_2\le 2.05 \\ 25A & Q_2\le 1.90 \\ 17A & Q_2>1.87 \end{cases}$
1.95A	$T = \begin{cases} ADA & 1.95<Q_2\le 2.10 \\ 25A & Q_2\le 1.95 \\ 17A & Q_2>2.10 \end{cases}$
2.00A	$T = \begin{cases} 21A & Q_2\le 2.00 \\ 12A & Q_2>2.00 \end{cases}$

Rank Estimators

In what follows, let $R_i(\theta)$ denote the signed ranks of $x_i-\theta$. That is, we rank $x_1-\theta, x_2-\theta, \ldots, x_n-\theta$ according to magnitude (but not sign) and $R_i(\theta)$ is the product of $\operatorname{sgn}(x_i-\theta)$ and rank of $(x_i-\theta)$. Further let us define $J(u) = \Phi^{-1}(\frac{1+u}{2})$, where Φ is the normal distribution function. The rank statistic with normal scores, (RN), is taken as the root, T, of the equation

$$J^*(T)-\bar{J} = 0$$

where

$$J^*(T) = n^{-1} \sum{}^* J\left(R_i(T)/n+1\right)$$

and

$$\bar{J} = n^{-1} \sum_{i=1}^{n} J(i/n+1).$$

The symbol, $\sum^*$, indicates the summation over all positive ranks. Roots were found by the method of bisection using a maximum of twelve iterations. If Φ is taken as uniform, the resulting rank statistic is asymptotically equivalent to the Hodges-Lehmann estimator (cf. Hodges-Lehmann (1963). Details of Johns (1974) and the skipped estimate may be found in the *PRS*.

3. MONTE-CARLO DETAILS

In addition to the long-tailed sampling situations investigated in the *PRS*, we also investigated short-tailed situations and

situations involving dependent data. Pseudo-random numbers were generated by a multiplicative congruential generator. Cycle time was constructed to be considerably larger than the number of observations needed for this study. We also checked for bias, variance, and serial correlation and the generator was suitable in these respects. (See the *PRS* for description and cautions in the use of these generators.) The output, scaled between 0 and 1, form the basis for all subsequent calculations since they approximate a uniformly distributed sample. Normal pseudo-random deviates were generated according to the well-known Box-Muller transform (cf. Paley and Wiener (1934), p. 146) and Cauchy pseudo-random deviates were generated by the tangent transform. These three basic distributions, or mixtures thereof, form the basis of all results reported in this work.

The basic sampling situations are summarized in Table II.

TABLE II

Sampling Situations for Independent Observations (n = 20). N/U is a Normal (0,1) Variate Divided by an Independent Uniform Variate

Distributions	Code
Uniform, mean 0, variance 1 (denoted U(0,1))	U
97% Uniform (0,1) + 3% Normal (0,1) contamination	U+.03N(0,1)
90% Uniform (0,1) + 10% Normal (0,1) contamination	U+.10N(0,1)
50% Uniform (0,1) + 50% Normal (0,1) contamination	U+.50N(0,1)
Normal, mean 0, variance 1	N(0,1)=N
99% Normal (0,1) + 1% Normal (0,9) contamination	N+.01N(0,9)
97.5% Normal (0,1) + 2.5% Normal (0,9) contamination	N+.025N(0,9)
95% Normal (0,1) + 5% Normal (0,9) contamination	N+.05N(0,9)
90% Normal (0,1) + 10% Normal (0,9) contamination	N+.10N(0,9)
75% Normal (0,1) + 25% Normal (0,9) contamination	N+.25N(0,9)
95% Normal (0,1) + 5% Normal (0,100) contamination	N+.05N(0,100)
90% Normal (0,1) + 10% Normal (0,100) contamination	N+.10N(0,100)
75% Normal (0,1) + 25% Normal (0,100) contamination	N+.25N(0,100)
97% Normal (0,1) + 3% Cauchy contamination	N+.03C
90% Normal (0,1) + 10% Cauchy contamination	N+.10C
50% Normal (0,1) + 50% Cauchy contamination	N+.50C
Cauchy, median = 0	C
90% Normal (0,1) + 10% N/U	N+.10N/U
75% Normal (0,1) + 25% N/U	N+.25N/U

In addition to the sets of independent observations discussed above, we generated observations according to a first order autoregressive scheme, $x_j = \rho x_{j-1} + \varepsilon_j$, $j = 1,2,\ldots,20$. The shock, ε_j , was chosen according to either N, U or C and the correlation coefficient, ρ, as either .2, .5 or .9. The initial value, x_0 , was chosen as 0. All of the 19 estimators discussed in section 2 were investigated in the 19 independent sampling situations. Only the non-adaptive estimators were studied in the 9 time-series sampling achemes.

The results we report are based on 1000 Monte-Carlo replications. The sample variances of the estimators were calculated and then scaled by a factor of 20 (the sample size) to make the results comparable to those in the *PRS*. Assuming approximate normality for the estimators, the variance of $20\times$ sample variance would be about $.7992\sigma_T^4$ where σ_T^2 is the true variance of the estimator, T. For example if $T = \bar{X}$, then σ_T^2 is $\frac{1}{20}$ and the variance of 20 times the sample variance is .001998. Thus one may expect about one decimal place of accuracy with correspondingly less significance as the value σ_T^2 climbs. We report two decimal places, in most cases, because all estimators were calculated over the same samples. Without attempting inferences outside these samples, the extra decimal place(s) are meaningful.

4. THE RESULTS

The main results of the Monte Carlo study are summarized in 2 tables. Table III is a table of sample variances (over the 1000 Monte-Carlo replications) for the 19 independent sampling schemes and 19 estimators of location. The symbol *** in Table III refers to a variance exceeding 100.00. Table IV is a table of biases and variances for the 9 sampling schemes involving first order autoregressive variates. None of the Hogg-type adaptive estimators are listed for the two-fold reason that the choice of Q_1 and Q_2 is predicated on independent observations and that the basic estimators — trimmed means, hampels, etc. — perform poorly. Some

TABLE III

Variances of 19 Estimators of Location Based on 1000 Monte Carlo Replications

NUMBER	CODE	U	U+.03N(0,1)	U+.10N(0,1)	U+.50N(0,1)	N(0,1)	N+.01N(0,9)	N+.025N(0,9)	N+.05N(0,9)	N+.10N(0,9)
1	10%	1.34	1.32	1.37	1.13	1.11	1.05	1.13	1.24	1.32
2	50%	2.48	2.43	2.51	1.82	1.55	1.49	1.54	1.67	1.72
3	M	1.01	.98	1.05	.94	1.05	1.04	1.24	1.45	1.82
4	OM	.61	.60	.67	.84	1.25	1.31	1.87	2.27	3.37
5	12A	1.82	1.97	1.85	1.42	1.14	1.13	1.26	1.36	1.46
6	17A	1.41	1.52	1.46	1.22	1.06	1.05	1.18	1.27	1.39
7	21A	1.20	1.30	1.26	1.11	1.05	1.01	1.16	1.24	1.38
8	25A	1.08	1.17	1.14	1.04	1.03	.98	1.13	1.23	1.38
9	ADA	1.19	1.31	1.28	1.11	1.06	1.02	1.17	1.28	1.43
10	HG1	.80	.80	.88	.96	1.11	1.06	1.19	1.28	1.41
11	HG2	1.39	1.37	1.41	1.19	1.16	1.12	1.21	1.31	1.37
12	1.81*A	1.11	1.21	1.17	1.08	1.05	1.01	1.18	1.26	1.41
13	1.90A	1.09	1.18	1.15	1.05	1.03	.99	1.15	1.25	1.40
14	1.95A	1.08	1.18	1.14	1.04	1.03	.98	1.14	1.24	1.39
15	2.00A	1.20	1.30	1.26	1.11	1.05	1.02	1.16	1.24	1.39
16	H/L	1.21	1.28	1.25	1.09	1.03	1.01	1.14	1.26	1.38
17	RN	.89	.94	.97	.94	1.01	.98	1.14	1.26	1.45
18	JØH	.88	.88	.93	.95	1.06	1.02	1.20	1.25	1.47
19	5T4	1.22	1.28	1.27	1.10	1.08	1.05	1.20	1.29	1.47

TABLE IV
Biased and Variances for 12 Estimators of Location with First-Order Autoregressive Data.
Note that *** is a Value Larger than 9999.99.

DISTRIBUTION ρ		NORMAL .2		NORMAL .5		NORMAL .9		CAUCHY .2		CAUCHY .5		CAUCHY .9		UNIFORM .2		UNIFORM .5		UNIFORM .9	
NUMBER	CODE	BIAS	VARIANCE	BIAS	VARIANCE	BIAS	VARIANCE	BIAS	VARIANCE	BIAS	VARIANCE	BIAS	VARIANCE	BIAS	VARIANCE	BIAS	VARIANCE	BIAS	VARIANCE
1	10%	-.01	1.63	.02	3.62	.02	62.0	.02	50.0	.25	7460.	-34.3	***	.00	1.87	-.01	4.10	.12	57.5
2	50%	-.00	2.01	.02	4.18	.02	65.3	.00	8.19	.01	65.7	-29.1	***	-.00	3.21	-.01	5.15	.12	60.7
3	M	-.01	1.57	.02	3.59	.03	61.0	.24	3742.	.98	***	-36.2	***	.00	1.50	-.01	3.68	.12	56.3
4	OM	-.01	1.79	.02	3.91	.04	60.1	.49	***	1.98	***	-41.2	***	.00	1.00	-.01	3.18	.11	54.9
5	12A	.00	1.75	.02	3.81	.02	63.9	.00	7.33	.02	50.9	-30.5	***	.00	2.43	-.02	4.63	.12	59.4
6	17A	-.01	1.66	.02	3.68	.02	62.9	.00	8.69	.02	61.8	-32.0	***	.00	1.99	-.01	4.26	.12	58.2
7	21A	-.01	1.62	.02	3.64	.02	62.4	.00	9.41	.02	68.0	-32.8	***	.00	1.76	-.01	4.06	.12	57.8
8	25A	-.01	1.59	.02	3.62	.02	61.9	.00	10.7	.03	79.6	-33.5	***	.00	1.62	-.01	3.89	.12	57.3
9	ADA	.00	1.64	.02	3.68	.02	62.4	.00	7.95	.02	55.6	-33.3	***	.00	1.79	-.01	4.06	.12	57.9
10	H/L	-.00	1.61	.01	3.64	.02	62.2	-.02	18.8	-.09	727.	-34.4	***	.00	1.75	-.01	4.04	.12	57.4
11	JØH	-.00	1.63	.02	3.74	.03	61.1	-.00	10.3	-.02	161.	-38.5	***	.01	1.32	-.01	3.68	.11	56.8
12	5T4	-.01	1.66	.02	3.70	.02	62.3	-.01	8.00	.03	59.8	-34.8	***	.01	1.74	-.01	4.05	.12	57.8

NUMBER	CODE	N+.25N(0,9)	N+.05N(0,100)	N+.10N(0,100)	N+.25N(0,100)	N+.03C	N+.10C	N+.50C	C	N+.10N/U	N+.25N/U
1	10%	2.02	1.22	2.16	9.12	1.13	1.22	2.70	8.34	1.23	1.80
2	50%	2.08	1.58	1.94	2.56	1.55	1.59	2.06	2.83	1.61	2.00
3	M	2.96	5.45	11.5	26.0	72.9	***	***	***	***	***
4	OM	5.80	17.3	38.1	85.3	***	***	***	***	***	***
5	12A	1.87	1.27	1.44	2.48	1.26	1.30	1.68	2.76	1.29	1.58
6	17A	1.89	1.18	1.37	2.82	1.17	1.22	1.72	3.21	1.21	1.51
7	21A	1.96	1.16	1.36	3.11	1.14	1.21	1.73	3.58	1.19	1.50
8	25A	2.07	1.15	1.39	3.60	1.11	1.18	1.89	3.93	1.18	1.53
9	ADA	1.98	1.19	1.42	2.67	1.15	1.24	1.74	3.01	1.23	1.57
10	HG1	2.14	1.28	1.74	4.26	1.16	1.30	1.88	3.30	1.31	1.69
11	HG2	1.93	1.29	1.66	2.80	1.20	1.30	1.84	2.83	1.30	1.66
12	1.81*A	1.97	1.19	1.42	2.59	1.15	1.24	1.69	2.84	1.22	1.54
13	1.90A	2.01	1.17	1.38	2.86	1.12	1.21	1.75	3.30	1.20	1.52
14	1.95A	2.01	1.16	1.38	3.01	1.12	1.20	1.74	3.31	1.20	1.52
15	2.00A	1.94	1.18	1.42	2.77	1.15	1.23	1.74	2.98	1.21	1.49
16	H/L	1.97	1.22	1.65	4.04	1.14	1.24	2.01	4.57	1.25	1.73
17	RN	2.24	1.28	1.95	5.93	1.12	1.28	2.46	6.76	1.31	1.93
18	JOH	2.19	1.24	1.52	4.38	1.16	1.27	1.73	3.25	1.28	1.62
19	ST4	1.98	1.24	1.42	3.73	1.20	1.25	1.66	3.00	1.26	1.54

preliminary calculations indicate that Hogg-type adaptors perform similarly under the time series alternatives.

We withhold discussion of these results to section 5.

5. DISCUSSION AND CONCLUSION

Table III contains many estimators with fairly comparable variance. In order to make the better estimators more apparent, we have constructed an additional table, Table V. Let T_i, $i = 1,\ldots,1000$ represent the 1000 observations of an estimator of location, and let

$$s_T^2 = \frac{1}{1000} \sum_{i=1}^{1000} (T_i - \bar{T})^2$$

be the sample variance. Assuming the T_i's are approximately

normal, then $\dfrac{1000s_T^2}{\sigma_T^2}$ is distributed approximately as a χ^2 with 999 degrees of freedom. Accordingly, the variance of s_T^2 is $.001998\sigma_T^4$ and the variance of $20s_T^2$ is $.7992\sigma_T^4$ as pointed out earlier. An estimate of the standard deviation of s_T^2 can be found by calculating $\sqrt{.001998}\, s_T^2 = .0447s_T^2$. Table V was constructed by calculating the estimated standard deviation for the estimator with the smallest variance.

If an estimator had a variance that fell within one standard deviation of the minimal variance, then it was replaced by a "0" in Table V. If the variance was more than one standard deviation but less than two, it was replaced with a "1" in Table V. Similarly for "2", "3" and "4". Finally if the variance was more than 5 standard deviations away from the minimal variance, it was simple replaced by "5". Thus Table V leaves a clear picture of the stronger estimators.

In terms of the 4 light-tailed alternatives the estimator of choice appears to be the outer mean. From Table V we observe that HG1, RN and JOH also perform creditably but significantly more poorly than OM.

We believe that the uniform is a highly artificial situation. The contamination of a normal by a distribution whose support is a bounded interval seems less so. While this very small selection of short-tailed alternatives is inadequate for sweeping committments to certain types of estimators, it is very suggestive of what is reasonable. It is clear that most estimators do very poorly in short-tailed situations. Some authors suggest that whenever short-tailed alternatives arise, we will be able to recognize them and having recognized them, use some appropriate estimator such as OM. We believe that this is really a very crude form of adaptive procedure based on intuition or some other form of non-statistical knowledge. It is therefore not really satisfactory, and we believe some more formal procedures such as Hogg's procedure are desirable.

Just as dramatic as the OM's good performance in the short-tailed situations is its bad performance in every long-tailed situation. Table VI also distinguishes several classes of strong performers in heavy-tailed situations. As a class, the hampels appear strong in spite of the appearance of 5's. The adaptive hampels also appear very strong with the added bonus of no 5's in long-tailed situations.

In the *PRS*, the concept of *deficiency* is introduced in order to provide a comparison of two estimators. The efficiency of an estimator under test relative to a standard estimator is defined by

$$\text{efficiency} = \frac{\text{variance of standard}}{\text{variance}}$$

and the deficiency by

$$\text{deficiency} = 1 - \text{efficiency}.$$

Notice that a negative deficiency means the estimator is more efficient than the standard. Thus deficiency is centered at 0 with negative meaning the standard is less efficient and positive meaning more efficient.

We feel the advantage of the zero reference point is outweighed by the following anomaly. In a sense, efficiencies of 2 and ½ mean the same thing (interchanging the roles of the standard estimator with the test estimator). The corresponding deficiencies of ½ and -1 are not symmetrically located about 0 and, on the intuitive level, not apparently related.

Rather than compute deficiency, we have computed the natural logarithm of the efficiency ratio which also has a 0 reference (for efficiency 1) and is symmetrical in the sense that $\ell n2 = -\ell n\frac{1}{2}$. The logarithm of the efficiency also has the advantage that in order to shift standards, only a simple subtraction is necessary. Table VI gives logarithms of efficiencies for a variety of the better performing estimators relative to 1.95A. To illustrate

TABLE V

Deviations of the Variances of 19 Estimates of Location

	U	U+.03N(0,1)	U+.10N(0,1)	U+.50N(0,1)	N(0,1)	N+.01N(0,9)	N+.025N(0,9)	N+.05N(0,9)	N+.10N(0,9)	N+.25N(0,9	N+.05N(0,100)	N+.10N(0,100)	N+.25N(0,100)	N+.03C	N+.1C	N+.5C	C	N+.1N/U	N+.25N/U
10%	5	5	5	4	2	1	0	0	0	1	1	5	5	0	0	5	5	0	4
50%	5	5	5	5	5	5	5	5	5	2	5	5	0	5	5	5	0	5	5
M	5	5	5	2	0	1	2	3	5	5	5	5	5	5	5	5	5	5	5
OM	0	0	0	0	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
12A	5	5	5	5	2	3	2	2	2	0	2	1	0	2	2	0	0	2	1
17A	5	5	5	5	1	1	0	0	1	0	0	0	3	1	0	0	3	0	0
21A	5	5	5	4	0	0	0	0	1	1	0	0	5	0	0	0	5	0	0
25A	5	5	5	4	0	0	0	0	1	2	0	0	5	0	0	2	5	0	0
ADA	5	5	5	4	1	0	0	0	1	1	0	0	1	0	1	0	2	0	1
HG1	5	5	5	3	2	1	1	0	1	3	2	5	5	0	2	2	4	2	2
HG2	5	5	5	5	3	3	1	1	0	0	2	4	1	2	2	2	0	2	2
1.81*A	5	5	5	4	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0
1.90A	5	5	5	4	0	0	0	0	1	1	0	0	3	0	0	0	4	0	0
1.95A	5	5	5	4	0	0	0	0	1	1	0	0	4	0	0	0	4	0	0
2.00A	5	5	5	4	0	0	0	0	1	0	0	0	2	0	0	0	1	0	0
H/L	5	5	5	4	0	0	0	0	1	1	1	4	5	0	1	4	5	1	3
RN	5	5	5	2	0	0	0	0	2	4	2	5	5	0	1	5	5	2	5
JOH	5	5	5	2	1	0	1	0	2	3	1	2	5	0	1	0	3	1	1
ST4	5	5	5	4	1	1	1	1	2	1	1	0	5	1	1	0	1	1	0

the change of standard, the log efficiency of 25A relative to 1.95A under Cauchy alternative is $-.172$ while the log efficiency of 1.81*A relative to 1.95A is $.153$. Thus the log efficiency of 25A relative to 1.81*A is just $-.172 - .153 = -.325$. Note that a negative log efficiency means the standard is less efficient than the test estimator.

In Test VI, 1.95A was used as a standard although it was not uniformly the best. However, we feel it adequately represents the class of adaptive hampels. For the moment, let us consider Table VI based on variances.

TABLE VI

Logarithm of the Relative Efficiency at 1.95A.

	17A	21A	25A	ADA	1.81*A	1.90A	2.00A	H/L	5T4
N(0,1)	-.029	.019	.000	-.029	-.019	.000	.010	.000	-.047
N+.01N(0,9)	-.069	-.030	.000	.040	-.030	-.010	-.040	-.030	-.069
N+.025N(0,9)	-.034	-.017	.009	-.026	.034	-.009	-.017	.000	-.051
N+.05N(0,9)	-.024	.000	.008	-.031	-.016	-.008	.000	-.016	-.062
N+.10N(0,9)	.000	.007	.007	-.028	-.014	-.007	.000	.007	-.056
N+.15N(0,9)	.013	.020	.007	-.020	.013	.007	.013	.007	-.032
N+.25N(0,9)	.062	.025	-.029	.015	.020	.000	.035	.020	.015
N+.05N(0,100)	-.017	.000	.009	-.026	-.026	-.008	-.017	.050	-.067
N+.10N(0,100)	.007	.015	-.007	-.029	-.029	.000	-.029	-.179	-.029
N+.25N(0,100)	.065	-.033	-.179	.120	.150	.051	.083	-.294	-.214
N+.03C	-.044	-.018	.009	-.026	-.026	.000	-.026	-.018	-.069
N+.10C	-.016	-.008	.017	-.033	-.033	-.008	-.025	-.033	-.041
N+.50C	.012	.006	-.083	.000	.029	-.006	.000	-.144	.047
C	.031	-.078	-.172	.095	.153	.003	.105	-.322	.098
N+.10N/U	-.008	.008	.017	-.025	-.017	.000	-.008	-.041	-.049
N+.25N/U	.007	.013	-.007	-.032	-.013	.000	.020	-.129	-.013

When compared to ADA, H/L and 5T4, 1.95A dominates in the sense that there are more negatives than positives. In these cases, the magnitude of the negative log efficiency is generally much greater than that of positive log efficiency. For example, the largest negative log efficiency for 5T4 is -.214 while the largest positive log efficiency is only .098. The only clear exception to this rule in these cases is the positive .120 for ADA. We can point out that relative to 1.81*A, ADA does not exhibit this behavior.

When compared to the hampels, the adaptive hampels do not usually dominate in terms of sign. That the adaptive hampels do not uniformly dominate is an entirely reasonable outcome. The adaptive estimators are, after all, dynamic averages of the non-adaptive versions. Therefore, we could not reasonably expect to achieve the very best behavior of the non-adaptives, but we should be able to get quite close. Moreover adaption allows us to avoid

the worst pitfalls. Thus, the worst case behavior still applies. For example, 25A has a positive log efficiency in 10 of the 16 cases. However, the maximum positive log efficiency is only .017 whereas the smallest negative log efficiency is -.179. For 22A the maximum positive log efficiency is .015 while the worst negative case is -.051, and so on.

To summarize, with adaptive hampels, one may lose a slight bit of efficiency in some cases, but gain a rather large amount in others. In this sense if hampels are good, adaptive hampels are better.

There is also some question in our minds whether Q_1 and Q_2 are the most suitable measures of heavy-tailedness. In particular, as Hogg has suggested in a private communication, although a sample may be drawn from a symmetric population the sample may have significant asymmetries. Thus an adaptive procedure accounting for such asymmetries should be able to improve procedures such as HG1 and HG2 as well as the adaptive hampels. Also we feel work needs to be done in adapting to light tails. A more sensitive measure of light-tailedness may improve the efficiencies there also. In regard to adaptive estimation, we must take exception to the *PRS*. Properly formulated, blatantly adaptive estimators perform at least as well as non-adaptors for sample sizes less than 40.

Finally we may address ourselves to Table IV and the time series alternative. A word on the design of the sampling situations in in order. For the first order autoregressive model

$$x_t = \rho x_{t-1} + \varepsilon_t , \qquad t = 1,2,\ldots,n$$

a negative value of ρ guarantees a spectrum dominated by high frequencies. In this circumstance observations will occur on both "sides" of the center. For positive values of ρ low frequencies dominate. Hence for positive ρ, particularly for ρ close to 1 the time series could make long excursions on one "side" of the center. Thus the difficulty in location estimation in a time

series is close to nonstationarity. Gastwirth and Rubin (1975) discuss robust estimation in precisely this situation. In particular, they derive efficiencies of several estimators relative to the mean M. Unfortunately, as Table IV clearly demonstrates, M is on the whole unsatisfactory. It is our assessment that no estimator presently fashionable is very satisfactory in the presence of positively correlated data.

A more complicated time series situation might have been to examine a time series which is contaminated by (possibly) uncorrelated observations. The general performance as seen in Table IV did not seem to warrant this investigation, however.

ACKNOWLEDGEMENTS

The work of Professor Wegman and the development of the computer routines were supported by the Air Force Office of Scientific Research under grant number AFOSR-75-2840.

BIBLIOGRAPHY

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972), *Robust Estimates of Location: Survey and Advances*, Princeton, N.J.: Princeton Univ. Press.

Bickel, P. J. (1975), One-step Huber estimates in the linear model, *J. Am. Statist. Assoc.*, *70*, pp. 428-434.

Box, G. E. P. and Muller, M. E. (1959), A note on the generation of random normal deviates, *Ann. Math. Statist.*, *29*, pp. 610-611.

Carroll, R. J. and Wegman, E. J. (1975), A Monte-Carlo study of robust estimators of location, Institute of Statistics Mimeo Series 1040, Univ. of North Carolina, Chapel Hill, N.C.

Gastwirth, J. and Rubin, H. (1975), The behavior of robust estimators on dependent data, *Ann. Statist.*, *3*, pp. 1070-1100.

Hampel, F. (1968), *Contributions to the Theory of Robust Estimation*, Ph.D. Dissertation, Univ. of California at Berkeley.

Hampel, F. (1974), The influence curve and its role in robust estimation, *J. Am. Statist. Assoc.*, *69*, pp. 383-393.

Hodges, J. L. and Lehmann, E. L. (1963), Estimates of location based on rank tests, *Ann. Math. Statist.*, *34*, pp. 598-611.

Hogg, R. V. (1974), Adaptive robust procedures: A partial review and some suggestions for future applications and theory, *J. Am. Statist. Assoc.*, *69*, pp. 909-923.

Huber, P. J. (1964), Robust estimation of a location parameter, *Am. Math. Statist.*, *35*, pp. 73-101.

Johns, M. V. Jr. (1974), Nonparametric estimators of location, *J. Am. Statist. Assoc.*, *69*, pp. 453-460.

Paley, R. E. A. C. and Wiener, N. (1934), *Fourier Transforms in the Complex Domain*, Providence, R.I.: American Mathematical Society.